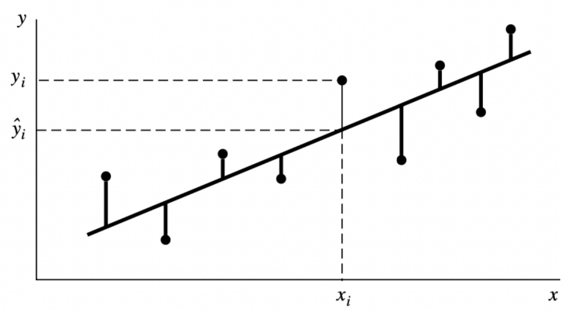
The Method of Least Squares

We use the method of least squares to estimate the parameters of any linear model. In the method of least squares, we fit a straight line to a set of data points. Suppose that we wish to fit the model

$$E(Y) = \beta_0 + \beta_1 x$$

to the set of data points shown in the following figure.



Fitting a straight line through a set of data points

We assume that $Y = \beta_0 + \beta_1 x + \epsilon$, where ϵ possesses some probability distribution with $E(\epsilon) = 0$. If $\hat{\beta}_0$ and $\hat{\beta}_1$ are estimators of the parameters β_0 and β_1, then $\hat{Y} = \hat{\beta}_0 + \hat{\beta}_1 x$ is clearly an estimator of $E(Y)$. We minimize the sum of squares of the vertical deviations from the fitted line. Thus if

$$\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_i$$

is the predicted value of the ith y value (when $x = x_i$), then the deviation (sometimes called the *error*) of the observed value of y_i from $\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_i$ is the difference $y_i - \hat{y}_i$, and the sum of squares of deviations to be minimized is

$$\text{SSE} = \sum_{i=1}^{n} (y_i - \hat{y}_i)^2 = \sum_{i=1}^{n} [y_i - (\hat{\beta}_0 + \hat{\beta}_1 x_i)]^2..$$

The quantity SSE is also called the *sum of squares for error*. If SSE possesses a minimum, it will occur for values of β_0 and β_1 that satisfy the equations, $\partial\text{SSE}/\partial\beta_0 = 0$ and $\partial\text{SSE}/\partial\beta_1 = 0$.

The equations $\partial\text{SSE}/\partial\beta_0 = 0$ and $\partial\text{SSE}/\partial\beta_1 = 0$ are called the *least-squares equations* for estimating the parameters of a line.

Taking the partial derivatives of SSE with respect to $\hat{\beta}_0$ and $\hat{\beta}_0$ and setting them equal to zero, we obtain

$$\hat{\beta}_1 = \frac{\sum_{i=1}^{n}(x_i - \overline{x})(y_i - \overline{y})}{\sum_{i=1}^{n}(x_i - \overline{x})^2} = \frac{\sum_{i=1}^{n} x_i y_i - \frac{1}{n}\sum_{i=1}^{n} x_i \sum_{i=1}^{n} y_i}{\sum_{i=1}^{n} x_i^2 - \frac{1}{n}\left(\sum_{i=1}^{n} x_i\right)^2},$$

$$\hat{\beta}_0 = \overline{y} - \hat{\beta}_1 \overline{x}.$$

It can be shown that the simultaneous solution for the two least-squares equations yields values of $\hat{\beta}_0$ and $\hat{\beta}_1$ that minimize SSE.

Least-Squares Estimators for the Simple Linear Regression Model

(1) $\hat{\beta}_1 = \frac{S_{xy}}{S_{xx}}$, where $S_{xy} = \sum_{i=1}^{n}(x_i - \overline{x})(y_i - \overline{y})$ and $S_{xx} = \sum_{i=1}^{n}(x_i - \overline{x})^2$.

(2) $\hat{\beta}_0 = \overline{y} - \hat{\beta}_1 \overline{x}$.

Question 1. Show that $\hat{\beta}_1$ and $\hat{\beta}_0$ are unbiased estimators for β_0 and β_1, respectively.

Question 2. Assume that the variance of the error term is σ^2. that is, $\text{Var}(\epsilon) = \sigma^2$. Compute the variances of the estimators $\hat{\beta}_0$ and $\hat{\beta}_1$. That is, compute $\text{Var}(\hat{\beta}_0)$ and $\text{Var}(\hat{\beta}_1)$.

Question 3. Usually the value of σ^2 is unknown, and we will need to make use of the sample observations to estimate σ^2. Can you find an unbiased estimator for the parameter σ^2?

Question 4.

(1) Use the method of least squares to fit a straight line to the $n = 5$ data points given in

Table 11.1 Data for Example 11.1

x	y
−2	0
−1	0
0	1
1	1
2	3

(2) Find the variances of the estimators $\hat{\beta}_0$ and $\hat{\beta}_1$.
(3) Estimate σ^2.